

UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Graph Representation Learning-Based Recommender Systems

by

Lei Sang

THESIS FOR
Doctor of Philosophy

Sydney, Australia

2020

Certificate of Original Authorship

I, Lei Sang declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution.

Also, I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with Hefei University of Technology, China.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 20/09/2020

ABSTRACT

Graph Representation Learning-Based Recommender Systems

by
Lei Sang

Personalized recommendation has been applied to many online services such as E-commerce and adverting. It facilitates users to discover a small set of relevant items, which meet their personalized interests, from many choices. Nowadays, various auxiliary information on users and items become increasingly available in online platforms, such as user demographics, social relations, and item knowledge. More recent evidences suggests that incorporating such auxiliary data with collaborative filtering can better capture the underlying and complex user-item relationships, and further achieve higher recommendation quality.

In this thesis, we focus on auxiliary data with graph structure, such as social networks and knowledge graphs(KG). For example, we can improve recommendation performance by mining social relationships between users, and also by using knowledge graphs to enhance the semantics of recommended items. Network representation learning aims to represent each vertex in a network (graph) as a low-dimensional vector while still preserving its structural information. Due to the availability of massive graph data in recommender systems, it is a promising approach to combine network representation learning with recommendation. Applying the learned graph features to recommender systems will effectively enhance the learning ability of the recommender systems and improve the accuracy and user satisfaction of the recommender systems. For network representation learning and its application in recommendation systems, the major contributions of this thesis are as follows:

(1) Attention-based Adversarial Autoencoder for Multi-scale Network Embedding. Existing Network representation methods usually adopt a one-size-fits-all approach when concerning multi-scale structure information, such as first- and second-order proximity of nodes, ignoring the fact that different scales play different roles in embedding learning. We propose an Attention-based Adversarial Autoencoder Network Embedding (AAANE) framework, which promotes the collaboration of different scales and lets them vote for robust representations.

(2) Multi-modal Multi-view Bayesian Semantic Embedding for Community Question Answering: Semantic embedding has demonstrated its value in latent representation learning of data, and can be effectively adopted for many applications. However, it is difficult to propose a joint learning framework for semantic embedding in Community Question Answer (CQA), because CQA data have multi-view and sparse properties. In this thesis, we propose a generic Multi-modal Multi-view Semantic Embedding (MMSE) framework via a Bayesian model for question answering.

(3) Context-Dependent Propagating-based Video Recommendation in Multi-

modal Heterogeneous Information Networks. Conventional approaches to video recommendation primarily focus on exploiting content features or simple user-video interactions to model the users' preferences. However these methods fail to model the complex video context interdependency, which is obscure/hidden in heterogeneous auxiliary data. In this paper, we propose a Context-Dependent Propagating Recommendation network (CDPRec) to obtain accurate video embedding and capture global context cues among videos in HINs. The CDPRec can iteratively propagate the contexts of a video along links in a graph-structured HIN and explore multiple types of dependencies among the surrounding video nodes.

(4) Knowledge Graph Enhanced Neural Collaborative Filtering. Existing neural collaborative filtering (NCF) recommendation methods suffer from severe sparsity problem. Knowledge Graph (KG), which commonly consists of fruitful connected facts about items, presents an unprecedented opportunity to alleviate the sparsity problem. However, NCF only methods can hardly model the high-order connectivity in KG, and ignores complex pairwise correlations between user/item embedding dimensions. To address these issues, we propose a novel Knowledge graph enhanced Neural Collaborative Recommendation (K-NCR) framework, which effectively combines user-item interaction information and auxiliary knowledge information for recommendation.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Min Xu for her selfless devotion to supporting and cultivating me in past three and a half years. She offered me a unique opportunity to study at UTS, where I was able to learn from so many talented colleagues. From the beginning of the scholarship application to the final thesis submission, Dr. Xu constantly gave me great support. Without her professional guidance and support, this work would not have been achieved.

I would also like to express my thanks to Dr. Shengsheng Qian for his generous guidance, encouragement and collaboration. He provided valuable suggestions to my research, including thinking and writing. I would like to pass my gratitude to my colleagues and friends in UTS: Tianrong Rao, Haimin Zhang, Lingxiang Wu, Zhongqin Wang, Zhiyuan Shi, Yukun Yang, Wanneng Wu, Ruiheng Zhang, Xiaoxu Li, Zhu Lin, Peng Zhang, Tao Shen etc. Thanks the UTS President's Scholarship and China Scholarship Council (CSC) Scholarship for funding my research.

Lastly, I want to show my special thanks to my family: my parents, my brother and my sister. Because of them, my every effort became meaningful. I especially thank my parents for their self-sacrifice and endless love. I am thankful to my beloved wife, Fengnan Ma, and my baby son, Xiayi Sang, for their love and encouragement which helps me finish this thesis.

List of Publications

Published Papers

1. **Lei Sang**, Min Xu, Shengsheng Qian, Xindong Wu. AAANE: Attention-based Adversarial Autoencoder for Multi-scale Network Embedding. PAKDD 2019.
2. **Lei Sang**, Min Xu, Shengsheng Qian, Xindong Wu. Multi-modal multi-view Bayesian semantic embedding for community question answering. Neurocomputing 334, 44-58, 2019.
3. **Lei Sang**, Min Xu, Shengsheng Qian, Xindong Wu. Context-Dependent Propagating based Video Recommendation in Multimodal Heterogeneous Networks. IEEE Transactions on Multimedia, 2020. DOI:10.1109/TMM.2020.3007330
4. **Lei Sang**, Min Xu, Shengsheng Qian, Xindong Wu. Knowledge Graph Enhanced Neural Collaborative Recommendation, Expert Systems with Applications, 2020. DOI: 10.1016/j.eswa.2020.113992
5. Lingxiang Wu, Min Xu, **Lei Sang**, Ting Yao, Tao Mei. Noise Augmented Double-stream Graph Convolutional Networks for Image Captioning, IEEE Transactions on Circuits and Systems for Video Technology. 2020

Others

1. **Lei Sang**, Min Xu, Shengsheng Qian, Xindong Wu. Knowledge Graph Enhanced Neural Collaborative Filtering with Residual Recurrent Network. submitted to Neurocomputing (Under Major Revision)
2. **Lei Sang**, Min Xu, Shengsheng Qian, Xindong Wu. Adversarial Optimisation for Heterogeneous Graph Neural Network based Recommendation, under submission to IEEE Transactions on Systems, Man, and Cybernetics: Systems.

Contents

Certificate	ii
Abstract	iii
Acknowledgments	v
List of Publications	vi
List of Figures	x
List of Tables	xii
1 Introduction	1
1.1 Background	1
1.2 Motivation & Contribution	3
2 Related Works	6
2.1 Introduction	6
2.2 Network Representation Learning	6
2.3 Graph-enhanced recommendation	8
2.3.1 Path-based methods	8
2.3.2 Embedding-based methods	9
2.4 Neural Collaborative Filtering	10
3 AAANE: Attention-based Adversarial Autoencoder for Multi-scale Network Embedding	11
3.1 Introduction	11
3.2 Preliminary	13
3.3 Method	15
3.3.1 An overview of the framework	15
3.3.2 Attention-based Autoencoder	15
3.3.3 Adversarial Learning	17
3.3.4 Training Procedure	17
3.4 Experiments	18

3.4.1	Datasets	18
3.4.2	Baselines and Experimental Settings	18
3.4.3	Multi-label Classification	20
3.4.4	Detailed Analysis of The Proposed Model	20
3.5	Conclusion	22
4	Bayesian Semantic Embedding for social-aware Community Question Answering	23
4.1	Introduction	23
4.2	The Proposed MMSE Model	26
4.2.1	Generative Process	26
4.2.2	Model Inference	29
4.2.3	Parameter Update	30
4.3	Question Answering	34
4.4	Experiments	36
4.4.1	Datasets and Evaluation	37
4.4.2	Baselines	38
4.4.3	Parameter Setting	39
4.4.4	Performance Evaluations and Comparisons	39
4.4.5	Impact of Parameter Values and Case Study	41
4.5	Experiment for expert finding	43
4.6	Conclusions	45
5	Context-Dependent Propagating-based Video Recommendation in Multimodal Heterogeneous Information Networks	46
5.1	Introduction	46
5.2	Preliminary	50
5.2.1	Framework	52
5.2.2	Construction of video graph from HIN	53
5.2.3	Multimodal video feature representation	55
5.2.4	Context-dependent propagating network	56
5.2.5	Video Prediction	59
5.3	Experiments	60

5.3.1	YouTube video dataset	60
5.3.2	Evaluation Metrics	60
5.3.3	Baselines	61
5.3.4	Experimental Results	64
5.3.5	Detailed analysis of the proposed model	65
5.3.6	Case study	67
5.3.7	Parameter sensitivity	68
5.4	Conclusion	69
6	Knowledge Graph Enhanced Neural Collaborative Recommendation	70
6.1	Introduction	70
6.2	PRELIMINARY	73
6.2.1	Problem Formulation	74
6.2.2	Neural collaborative filtering framework	75
6.3	Knowledge Enhanced Neural Collaborative Model	77
6.3.1	Framework	77
6.3.2	Knowledge propagation based item modelling	77
6.3.3	Attention-based user modelling	80
6.3.4	Convolutional NCF	82
6.3.5	Learning Algorithm	84
6.4	Experiment	85
6.4.1	Datasets	85
6.4.2	Baselines	86
6.4.3	Experiment Settings	88
6.4.4	Overall Performance	89
6.4.5	Parameter Sensitivity	91
6.5	Conclusion	93
7	Conclusion and Future Work	94
7.1	Conclusion	94
7.2	Future Work	95
	Bibliography	96

List of Figures

1.1	Graph structure data used in recommendation.	2
2.1	Graph structure of the Zachary Karate Club social network (left) and Two-dimensional visualization of node embeddings generated from this graph using the DeepWalk method (right). Reprinted from [1].	7
3.1	Illustration of the multi-scale network with three scales.	12
3.2	The architecture of AAANE. To learn the low dimension of node representation, we employ autoencoder with non-linear activation function to capture complex, non-linear projections from the inputs to the outputs. The We first introduce an attention mechanism to the autoencoder for learning the weights of structure information with different scales. Finally, the adversarial learning component acts as regularization for the autoencoder, by matching the aggregated posterior, which helps enhance the robustness of the representationThe.	14
3.3	Comparison of performances on each individual scale and the average weights of scales. Scales with better performances usually attract more attentions from nodes.	21
3.4	Parameter sensitivity of dimension d and scale size k	21
4.1	An example of multi-modal multi-view semantic embedding for CQA data. (i) The two modalities, Q/A pairs and corresponding users are explored jointly. (ii) Semantic embedding was modeled from both local and global views. Single-view based embedding method may lead to a biased result, such as local embedding for answer 2 and global topical embedding for answer 3.	24
4.2	The flowchart of the proposed multi-modal multi-view semantic embedding model.	25
4.3	The graphical representation of the proposed multimodal multi-view Semantic Embedding model (MMSE). Each document d contains a single social network user, one answer posted by the user and question related to the answer. Embeddings are observed variables. .	26

4.4	Distribution of number of answers given per user. A small fraction of users answer a lot of questions while many users answer a few number of questions.	37
4.5	Effect of varying the balance weight c	42
5.1	(a) An example of content-based recommendation. Only the video content feature is used to model user preference. (b) Collaborative filtering uses simple user-video interaction for video recommendation. (c) Our general idea of introducing a multimodal HIN to model user historical information. We try to encode global video context and multimodal video content feature into the comprehensive video embedding for recommendation purposes. . . .	47
5.2	Network schemas of a heterogeneous information network used for our social video recommendation scenario.	50
5.3	The overall framework CDPRec.	53
5.4	Performance comparison of different methods for cold-start recommendation YouTube datasets. y-axis denotes the improvement ratio over CMF.	65
5.5	Performance change of CDPRec when meta-paths are gradually added.	66
5.6	Visualization of three paths with relevance probabilities for a user. . .	68
5.7	Parameter sensitivity of CDPRec.	69
6.1	Illustration of KG-enhanced movie recommender systems. The knowledge graph can provides extra user-item connectivity information, which are useful for alleviating the sparsity problem of target recommendation task.	71
6.2	Neural Collaborative Filtering (NCF): MF as a shallow neural network model.	75
6.3	Overall architecture of K-NCR. There are three components in the framework: (1) an knowledge aggregating-based Item modelling (Left) (2) attention-based User Modelling (Low right) and (3) Convolutional NCF for item/user interaction modelling (Upper right).	76
6.4	The long-tail distribution of Movielens-20M dataset.	82
6.5	Hidden CNN layers with embedding size 64.	83
6.6	Computational cost comparison.	88
6.7	Performance of K-NCR <i>w.r.t.</i> different numbers of feature maps per convolutional layer (denoted by C) in each epoch.	91
6.8	Impact of parameter α	91

List of Tables

3.1	Accuracy (%) of node classification on Wiki.	19
3.2	Accuracy (%) of node classification on Cora.	19
3.3	Accuracy (%) of node classification on Citeser.	19
4.1	Key Notations of Our Proposed MMSE Model	27
4.2	Statistics of the two CQA datasets	36
4.3	Evaluation Results on Quora and Zhihu Data	40
4.4	Retrieval Results for “What should I know before traveling to Sydney? ”	41
4.5	Performance comparison of MMSE over different dimensionality of word embedding	42
4.6	Experimental results on nDCG with different proportions of Quora data for training	44
4.7	Experimental results on Precision@1 with different proportions of Quora data for training	44
5.1	Notations and explanations	51
5.2	Statistics of YouTube dataset.	61
5.3	Experimental Results on YouTube Data. We indicate the data source for each model. (1) “CF” denotes the user-item interaction for Collaborative Filtering purpose. (2) “CT” denotes the content feature of the video. (3) “HIN” denotes the Heterogeneous Information Network that consists by users and attribute information of videos. (4) “Social” denotes the social relationship among users.	62
5.4	The results w.r.t. different layer number.	67
6.1	Notations and explanations	74
6.2	Basic statistics settings for the three datasets	85
6.3	The results of AUC and $F1$ in CTR prediction.	87

6.4	Results of AUC on MovieLens-20M in CTR prediction with different ratios of training set r	90
6.5	Effect of embedding propagation layer numbers (L).	92